

```

FEATURES
  source
    Location/Qualifiers
      1..4015
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /tissue_type="kidney"
        /map="5p"
        /dev_stage="embryo"
        /chromosome="5"
      gene
        1..4015
        /gene="hprt"
      CDS
        56..3454
        /gene="hprt"
        /codon_start=1
        /product="telomerase reverse transcriptase"
        /db_xref="PID:G2330017"
        /translation="NPRAPRCRAVRSLLRSHRYEVLPLATFVRRIGPGQGWRIIVORGDDP

```

AAAPRALVAOCILVCVPMWDRPAPSPROVSCLEKELVARVLORLCERGAKNVLAFFGA
LLDARGCGPEAETTSYRSYLPTNTYTDALAGSGANGLLLRVGGDDVLHLLARLCAFLV
LVAPSAIYVCGGPLYQLGAATQAPPPHAGSPRRRLUGCCERAWNHSVREAGVPLGLFA
PGARRGGSASRLSLPRLPRRGAPEPTPVGQSWAHPGRTGRSPDSRAGFCVYLSA
RPAEERATSLGALSGLRSHPSVGRQHAGPPSTSRPPRWDTPCPVPVYAEKHFLLYS
SGDKEOLRPSFLLSLRPSLTHGARLVETIFLGRPMWPGTPRLPLRPLORYWOMRPL
FLELGNHACQPYGVLLKTHCPLRAVTPAAGVCAREKPGQSVAAPEEDTDRLVQ
LLRQHSPPWQYGVFRACLRRLVPPGLWGSNHNRRFLRNKKFISLGKNAKLSLQEL
TWKMSVDCAWLRSPGVGPVAAEHLREELAKFLHMLSVTVVVELLSFFVTET
TFQNRFLFKRSVMSKLSIGIRQHILKRLQRLSEAEVROHREAPALITSLRRI
KPDGLRIYVMDYVVGARTFRREKRAERLTSRKALFSLVLYNEARARPGLLGASVLG
LDDIHRARWTFVLVRAQDPPPELYFVKVDVTGAYDTIPDRLTEVASIIPKONTVC
VRYAVVQKAAGHVRAKAFKSHVSTLDLQPYMROFYAHLETSPLDADAVIEGSSSL
NEASGGLFVFLRFMCHAVIRKSTVQCGIPQGSILSLSCISLCYGDMEKFLFAG
IRRDGLLRLVDLFTVPLHTHAKEFLRLTVRGVPEYGVVYLNKTVYVNFVYDEAL
GGTAFVQMPAGHFFPMWGLLDTLREVSQDSYARTSRASLTFNRFGKAGNMRR
KLFVGLRLKCHSLFLDQVNSLQTVNIVYKLLLOAYRFRHACVLQPLFFHOQWKNPT
FFLRVISDTASLQVSIILKAKNAGMSLGAAGPLPSEAVQWLCHQAFLLKLRHRTV
YVPLLSLRPAQQLSRKLPGLTILTALEAANPALPSDFKILTD

COUNT 663 a 1363 c 1275 g 714 t

Query Match 100.0%; Score 4015; DB 21; Length 4015;

Best Local Similarity 100.0%; Pred. No. 0.00e+00;

Matches 4015; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db	1	GCAGCGTCCGCTGCTGCGACGTGGGAAGCCGTGCCCGCCGACCCCGCGATGCC	60
Qy	1	GCAGCGTCCGCTGCTGCGACGTGGGAAGCCGTGCCCGCCGACCCCGCGATGCC	60
Db	61	CGGGCTCCCGCTGCGAGCGCTGCGCTCCCTGCTGCGAGCCACTACCGGGAGGTGCT	120
Qy	61	CGGGCTCCCGCTGCGAGCGCTGCGCTCCCTGCTGCGAGCCACTACCGGGAGGTGCT	120
Db	121	GCCCTGCGCAGCTGCTGCGCGCGCTGCGGCGCCGAGGCTGGCGGCTGCGAGCGGG	180
Qy	121	GCCCTGCGCAGCTGCTGCGCGCGCTGCGGCGCCGAGGCTGGCGGCTGCGAGCGGG	180
Db	181	GGACCCGCGGCTTTCCGCGCGCTGCTGCGCGCGCTGCTGCTGCTGCTGCGAGCGC	240
Qy	181	GGACCCGCGGCTTTCCGCGCGCTGCTGCGCGCGCTGCTGCTGCTGCTGCGAGCGC	240
Db	241	ACGGCG	300
Qy	241	ACGGCG	300
Db	301	CCGAGTCTGAGAGGCTGCTGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	360
Qy	301	CCGAGTCTGAGAGGCTGCTGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	360
Db	361	GCTGCTGGACGGCG	420
Qy	361	GCTGCTGGACGGCG	420
Db	421	CTGCG	480
Qy	421	CTGCG	480
Db	481	CCGGTGGCGGACGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	540
Qy	481	CCGGTGGCGGACGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	540
Db	541	GGCTCCGAGCTGCGCTTACAGAGTGTGGGGCGCGCGCGCTGCTGCTGCTGCTGCTG	600
Qy	541	GGCTCCGAGCTGCGCTTACAGAGTGTGGGGCGCGCGCGCTGCTGCTGCTGCTGCTG	600
Db	601	TCAGGCG	660
Qy	601	TCAGGCG	660
Db	661	CTGGAACCATAGCTCAGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	720
Qy	661	CTGGAACCATAGCTCAGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	720

Db	721	GAGGCGCGGGGCGAGTCCAGCGCGAAGTCTGCCGTTGCCCAAGAGGCCAGCGCTGGCGC	780
Qy	721	GAGGCGCGGGGCGAGTCCAGCGCGAAGTCTGCCGTTGCCCAAGAGGCCAGCGCTGGCGC	780
Db	781	TGCCCTCAGCGGAGCGAGCGCGGTTGGCAGGGGTCTTGGGCGCCACCCGGGCGAGAC	840
Qy	781	TGCCCTCAGCGGAGCGAGCGCGGTTGGCAGGGGTCTTGGGCGCCACCCGGGCGAGAC	840
Db	841	CGGTGGACCGAGTACCGCTGCTTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT	900
Qy	841	CGGTGGACCGAGTACCGCTGCTTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT	900
Db	901	CACCTCTTTTGGAGGTGCGCTCTCTGGCAGCGGCACCTCCACCCATCGTGGGCGCGCA	960
Qy	901	CACCTCTTTTGGAGGTGCGCTCTCTGGCAGCGGCACCTCCACCCATCGTGGGCGCGCA	960
Db	961	GCACACGCGGCG	1020
Qy	961	GCACACGCGGCG	1020
Db	1021	CCCGGTGTACGCGGAGCAAGCACTTCTCTACTCTCTCAGGCGCAAGAGGACGCTGG	1080
Qy	1021	CCCGGTGTACGCGGAGCAAGCACTTCTCTACTCTCTCAGGCGCAAGAGGACGCTGG	1080
Db	1081	GCCCTCTCTTCTACTACTACTCTCTGTAGGCCAGCGCTGCTGCGGCTCGGAGGCTCGTGA	1140
Qy	1081	GCCCTCTCTTCTACTACTACTCTCTGTAGGCCAGCGCTGCTGCGGCTCGGAGGCTCGTGA	1140
Db	1141	GACCATCTTTCTGGGTCCAGGCGCTGATGCCAGGACTCCCGCGAGGTTGCCCGGCT	1200
Qy	1141	GACCATCTTTCTGGGTCCAGGCGCTGATGCCAGGACTCCCGCGAGGTTGCCCGGCT	1200
Db	1201	GCCCGAGCGCTACTGGCAATGCGGCGCTGTTCTGAGGCTGTTGGAAACACGCGCA	1260
Qy	1201	GCCCGAGCGCTACTGGCAATGCGGCGCTGTTCTGAGGCTGTTGGAAACACGCGCA	1260
Db	1261	GTGCGGCTTACGGGCTGCTCTCAAGACGACTGCGCGCTGCGAGTGGCGTCAACCGGAC	1320
Qy	1261	GTGCGGCTTACGGGCTGCTCTCAAGACGACTGCGCGCTGCGAGTGGCGTCAACCGGAC	1320
Db	1321	AGCGGCTGCTGTGCGCGGAGAGCGCGCGCTGTGGGCGCGCGCGCGCGCGCGCGCG	1380
Qy	1321	AGCGGCTGCTGTGCGCGGAGAGCGCGCGCTGTGGGCGCGCGCGCGCGCGCGCGCG	1380
Db	1381	CACAGACCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	1440
Qy	1381	CACAGACCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	1440
Db	1441	CGGCTTCTGCGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	1500
Qy	1441	CGGCTTCTGCGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	1500
Db	1501	CAACGAGCGCGCTTCTCTCAGGAGACACCAAGATTCATCTCCCTGGGGAAGCATGCCAA	1560
Qy	1501	CAACGAGCGCGCTTCTCTCAGGAGACACCAAGATTCATCTCCCTGGGGAAGCATGCCAA	1560
Db	1561	GCTCTGCTGAGAGGCTGAGTGGAGATGAGGCTGCGGAGCTGCGCTGGCTGGCGAG	1620
Qy	1561	GCTCTGCTGAGAGGCTGAGTGGAGATGAGGCTGCGGAGCTGCGCTGGCTGGCGAG	1620
Db	1621	GAGCCGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	1680
Qy	1621	GAGCCGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	1680
Db	1681	CAAGTCTCTGACGTGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT	1740
Qy	1681	CAAGTCTCTGACGTGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT	1740
Db	1741	TGTACGAGAGACGCTTCAAGAGACAGGCTCTTTTCTACCGGAGAGTGTCTGGAG	1800
Qy	1741	TGTACGAGAGACGCTTCAAGAGACAGGCTCTTTTCTACCGGAGAGTGTCTGGAG	1800

QY 3961 GAGGTGCTGTGGAGTAAACTGAATATATAGTATTTTCAGTTTGAAGAAAAA 4015

RESULT 2
 LOCUS AF018167 4027 bp mRNA PRI 29-AUG-1997
 DEFINITION Homo sapiens telomerase catalytic subunit (hEST2) mRNA, complete cds.
 ACCESSION AF018167
 NID 92347128
 KEYWORDS human.
 SOURCE Homo sapiens
 ORGANISM Homo sapiens
 Eukaryotes; mitochondrial eukaryotes; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 4027)
 AUTHORS Meyerson, M., Counter, C.M., Eaton, E.N., Ellisen, L.W., Steiner, P., Caddie, S.D., Ziaugra, L., Beijersbergen, R.L., Davidoff, M.J., Liu, Q., Bacchetti, S., Haber, D.A. and Weinberg, R.A.
 hEST2, the putative human telomerase catalytic subunit gene, is up-regulated in tumor cells and during immortalization
 Cell 90 (4), 785-795 (1997)
 97433088

2 (bases 1 to 4027)
 AUTHORS Meyerson, M., Counter, C.M., Eaton, E.N., Ellisen, L.W., Steiner, P., Caddie, S.D., Ziaugra, L., Beijersbergen, R.L., Davidoff, M.J., Liu, Q., Bacchetti, S., Haber, D.A. and Weinberg, R.A.
 Direct Submission
 Submitted (11-AUG-1997) Whitehead Institute for Biomedical Research, Massachusetts Institute of Technology, Nine Cambridge Center, Cambridge, MA 02142, USA
 Location/Qualifiers
 1. 4027
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="5"
 /map="5p15.33"
 1. 4027
 /note="human Ever Shorter Telomeres 2"
 /gene="hEST2"
 57..3455
 /gene="hEST2"
 /note="member of reverse transcriptase class"
 /codon_start=1
 /db_xref="PDB:2347129"
 /translation="MPAPRCRAVRSLLRSHYREVLPATFVRRLGPGQRLVQRGDP
 AFRALVACLVCPWDARPPAPSPFRQVSCLELVARVQLRCERGANVLAFAFV
 LLDGARGGPEAFETTSVRSYLPNTVDALRGAGWGLLRVGGDDVLVHLLCALFV
 LVAPSCAYQVGPPLQLGAATQAPPPHAGSPRRRLGCRANVHVSREAGVPLGPA
 PGARRGGASRLPLPKPRRGAAPERTPVQGSWAHPGRTGSPDRGFVSPAP
 RPAEATSLGALSGTRHSPSVGROHAGPPTSRRPWPMTPCPPVATKHFYLS
 SGDEQRLRSFLLSSLPSTLTGARRLVETILGSRPWPMTPRLPRLPORYQWRPL
 FLELLGHAQCPGVLLKTHCPRLAAVTPAAGVAREKPGQSVAAPEEEDTDPRRLVQ
 LKHQSPQVQTVFACRLRVPLGLWSRHRERFLRNKFIISLGHAKLSLQEL
 TWKMSVGGCAKRSFVCPAAHREELILAKFLHMLMSVYVVELLSFFYVET
 TFQKRLFFYKRSWKLSQIGIRKHLKRVOLRELSEARQHREARPAALLSLRFL
 PKPDGLRVINQDYVVGARTFREKRAERLTSRVKAFLSVLNYREARPPGLGASVIG
 LDDIHRARTFVLVRAQDPPELYFVVDVTGAYDITPDRLTEVIASLIKQNYVC
 VRYIVQAAHGHKRAFKSHVSTLDLQPMRQFVAHQETSPURDAVVIQSSSL
 NEASSGLVDFLFMCHEAVRIRGSKYVQCGIQGSLSLCLCYGDMENKFLAG
 IRDGLLLRLDLDLVTPLHTLRTLEQSDYSYARTSTRASLTNRGPKAGNNMR
 GGTAFFQMPAHGLFPCWGLLDLRTLEQSDYSYARTSTRASLTNRGPKAGNNMR
 KLGLVLRKCHSLFLDQVNSLOTVCNINIKLLQNYRPHACVLOLPFHQWKNPT
 FFLRVLSDRASLTSLIKAKNAGSLAGAGLPFSEAVQWLCHEQAFLLKLRHRT
 YVPLGLSLTAQLRSLPGTTLTALRAANPALPSPDFKTLID"
 674 a 1361 c 1277 g 715 t

BASE COUNT 674 a 1361 c 1277 g 715 t

ORIGIN

Query Match 99.7%; Score 4001; DB 21; Length 4027;
 Best Local Similarity 100.0%; Pred. No. 0.00e+00;
 Matches 4003; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

12 GTCCTGCTGGCAGCTGGGAGAGCCCTGGCCCCCGGCCACCCCGCGATGCCGCGGCTCCC 71
 11 GTCCTGCTGGCAGCTGGGAGAGCCCTGGCCCCCGGCCACCCCGCGATGCCGCGGCTCCC 70
 72 CGCTGCCGAGCGCTGCTCCCTGCTGCGAGACCACTACCGCAGAGTGTCTCCGCTGGCC 131
 71 CGCTGCCGAGCGCTGCTCCCTGCTGCGAGACCACTACCGCAGAGTGTCTCCGCTGGCC 130
 132 ACCTTCGTGCGGCGCTTGGGCGCCCGAGGGTGGCGGTGTGTGACAGCGGGGACCCGCG 191
 131 ACCTTCGTGCGGCGCTTGGGCGCCCGAGGGTGGCGGTGTGTGACAGCGGGGACCCGCG 190
 192 GCTTTCGCGCGCTGTGTGGCCAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 251
 191 GCTTTCGCGCGCTGTGTGGCCAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 250
 252 CCCGCCCGCCCTCTTCCGCGCAGTGTCTGCTGCTGAAGAGCTGTGTGTGTGTGTGTGTGT 311
 251 CCCGCCCGCCCTCTTCCGCGCAGTGTCTGCTGCTGAAGAGCTGTGTGTGTGTGTGTGTGT 310
 312 CAGAGGCTGTGCGAGCGCGCGCGAAGAACTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 371
 311 CAGAGGCTGTGCGAGCGCGCGCGAAGAACTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 370
 372 GGGCGCGCGGCG 431
 371 GGGCGCGCGGCG 430
 432 ACCTGACCGACGACGCTGT 491
 431 ACCTGACCGACGACGCTGT 490
 492 GACGACGCTGT 551
 491 GACGACGCTGT 550
 552 TGCCTGTACAGAGT 611
 551 TGCCTGTACAGAGT 610
 612 CCCCG 671
 611 CCCCG 670
 672 AGCTGTAGGAGGCG 731
 671 AGCTGTAGGAGGCG 730
 732 GGCAGTGTACAGAGT 791
 731 GGCAGTGTACAGAGT 790
 792 CCGAGAGGAGCG 851
 791 CCGAGAGGAGCG 850
 852 AGTACCGCTGT 911
 851 AGTACCGCTGT 910
 912 GAGGT 971
 911 GAGGT 970
 972 GGGCG 1031
 971 GGGCG 1030
 1032 GCGAGAGGAGCG 1091
 1031 GCGAGAGGAGCG 1090
 1092 CTACTCAGCTCTGTAGGCG 1151

QY 190 GGCTTTCCGGCGCTGGTGGCCAGTGCCTGGTGTGGGTGCCCTGGGACGACGCGCGCC 249
Db 1341 CCCCGCGCCCTCTCTCCGCCAGTGTCTCTCCCTGAAGAGCTGCTGGCCCGAGTGTCT 1400
QY 250 CCCCGCGCCCTCTCTCCGCCAGTGTCTCTCCCTGAAGAGCTGCTGGCCCGAGTGTCT 309
Db 1401 GCAGAGGCTGTGGGAGCGGGCGCGAAGACGTGTGGCCCTTGGCTTCCGGTGTCTGGA 1460
QY 310 GCAGAGGCTGTGGGAGCGGGCGCGAAGACGTGTGGCCCTTGGCTTCCGGTGTCTGGA 369
Db 1461 CGGGGCGCGCGGGCGCCCGCCAGGCTTACACACAGGCTGCGAGCTACCTGCCCAA 1520
QY 370 CGGGGCGCGCGGGCGCCCGCCAGGCTTACACACAGGCTGCGAGCTACCTGCCCAA 429
Db 1521 CACGCTGACCGACGCTGTGGGGGAGCGGGCGCTGGGGCTGTGTGGCGCGCTGGG 1580
QY 430 CACGCTGACCGACGCTGTGGGGGAGCGGGCGCTGGGGCTGTGTGGCGCGCTGGG 489
Db 1581 CGACGAGCTGTGTTTACCTGTGTGACGCTGCGGGCTCTTTGTGTGTGTGTGTGTGT 1640
QY 490 CGACGAGCTGTGTTTACCTGTGTGACGCTGCGGGCTCTTTGTGTGTGTGTGTGTGT 549
Db 1641 CTGCGCTACCAAGTGTGGGGCGCGCTGTACCAAGCTGCGGGCTGCCACTCAGGCGCG 1700
QY 550 CTGCGCTACCAAGTGTGGGGCGCGCTGTACCAAGCTGCGGGCTGCCACTCAGGCGCG 609
Db 1701 GCCCGCGCACAGCTAGTGGACCCCGAAGCGCTGTGGATGCGAAGCGCTTGGAACTA 1760
QY 610 GCCCGCGCACAGCTAGTGGACCCCGAAGCGCTGTGGATGCGAAGCGCTTGGAACTA 669
Db 1761 TAGCGCTAGGGAGCGGGTCCCTTGGGCTTCCAGCCCGGGTGTGGAGGCGCGG 1820
QY 670 TAGCGCTAGGGAGCGGGTCCCTTGGGCTTCCAGCCCGGGTGTGGAGGCGCGG 729
Db 1821 GGGCAGTGCAGCGCAAGTGTGCGCTTGCCTAAGAGGCGCGAGCGCTGCCCTGA 1880
QY 730 GGGCAGTGCAGCGCAAGTGTGCGCTTGCCTAAGAGGCGCGAGCGCTGCCCTGA 789
Db 1881 GCCGAGCGGACGCCGTTGGGAGGGTCTTGGGCCACCGCGGAGGAGCGGTGAGCC 1940
QY 790 GCCGAGCGGACGCCGTTGGGAGGGTCTTGGGCCACCGCGGAGGAGCGGTGAGCC 849
Db 1941 GAGTGACCTGTTTCTGT 2000
QY 850 GAGTGACCTGTTTCTGT 909
Db 2001 GAGGCTGCGCTCTGTGGACGCGGCTCTCCACCTCCAGCTGCGGGCGCGAGCACGCG 2060
QY 910 GAGGCTGCGCTCTGTGGACGCGGCTCTCCACCTCCAGCTGCGGGCGCGAGCACGCG 969
Db 2061 GGGCCCCCATCCACATCGGGCCACCACTGCTTGGGACAGCGCTTGTCCCGGTTGA 2120
QY 970 GGGCCCCCATCCACATCGGGCCACCACTGCTTGGGACAGCGCTTGTCCCGGTTGA 1029
Db 2121 CGCCGAGACCAAGCACTTCTCTACTCTCTCAGCGGACCAAGGAGCGGTGGCGCTCTT 2180
QY 1030 CGCCGAGACCAAGCACTTCTCTACTCTCTCAGCGGACCAAGGAGCGGTGGCGCTCTT 1089
Db 2181 CTTACTCAGCTCTCTGAGGCCAGCTGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2240
QY 1090 CTTACTCAGCTCTCTGAGGCCAGCTGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1149
Db 2241 TCTGGTTCCAGGCTTGT 2300
QY 1150 TCTGGTTCCAGGCTTGT 1209
Db 2301 CTACTGGCAAAATCGGCGCTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2360
QY 1210 CTACTGGCAAAATCGGCGCTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1269
Db 2361 CGGGGTGCTCTCAAGACGCACTGCGGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2420
QY 1270 CGGGGTGCTCTCAAGACGCACTGCGGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1329

Db 2421 CTGTGCCCCGGGAAGCCCCCAGGGCTCTGTGGGGCCCCCGAGGAGGAGACACAGACCC 2480
QY 1330 CTGTGCCCCGGGAAGCCCCCAGGGCTCTGTGGGGCCCCCGAGGAGGAGACACAGACCC 1389
Db 2481 CCGTGGCTGGTGGAGCTGTCTCCGCCAGCAGCAGAGCCCTTGGCAGGTGTACGGCTTCGT 2540
QY 1390 CCGTGGCTGGTGGAGCTGTCTCCGCCAGCAGCAGAGCCCTTGGCAGGTGTACGGCTTCGT 1449
Db 2541 CGGGGCTGCTGCGCGGCTGTGGCCCCCAGGGCTCTGGGGCTCCAGGCAACAGAACG 2600
QY 1450 CGGGGCTGCTGCGCGGCTGTGGCCCCCAGGGCTCTGGGGCTCCAGGCAACAGAACG 1509
Db 2601 CCGGTTCTCAGGAACACCAAGAGTTTCTCTCCCTGGGAAGCATGCCAAGCTTCTGCT 2660
QY 1510 CCGGTTCTCAGGAACACCAAGAGTTTCTCTCCCTGGGAAGCATGCCAAGCTTCTGCT 1569
Db 2661 GCAGGAGCTGACGTGGAAGATGAGCGTGGGGGCTGCGCTTGGCTGCGAGGAGCCAGG 2720
QY 1570 GCAGGAGCTGACGTGGAAGATGAGCGTGGGGGCTGCGCTTGGCTGCGAGGAGCCAGG 1629
Db 2721 GGTGGCTGTCTCCGCGCGCAGAGCAGCTGCGTGAGGAGATCCTGGCCAAAGTTCT 2780
QY 1630 GGTGGCTGTCTCCGCGCGCAGAGCAGCTGCGTGAGGAGATCCTGGCCAAAGTTCT 1689
Db 2781 GCAGTGGCTGATGAGTGTACGTGCGAGCTGCTCAGGCTTCTTTTATGTACAGGA 2840
QY 1690 GCAGTGGCTGATGAGTGTACGTGCGAGCTGCTCAGGCTTCTTTTATGTACAGGA 1749
Db 2841 GACCACGTTTCAAAAGAACAGGCTCTTTTCTACCGAAGAGTGTCTGGAGCAAGTTGCA 2900
QY 1750 GACCACGTTTCAAAAGAACAGGCTCTTTTCTACCGAAGAGTGTCTGGAGCAAGTTGCA 1809
Db 2901 AAGCATTTGGAATCAGACAGCACTTGAAGAGGCTGACCTGGGAGGAGTGTGCGAAGCAGA 2960
QY 1810 AAGCATTTGGAATCAGACAGCACTTGAAGAGGCTGACCTGGGAGGAGTGTGCGAAGCAGA 1869
Db 2961 GGTGAGCAGCATGCGGAAGCCAGGCGCGCTGCTGACGTCCAGACTCCGCTTTCATCCC 3020
QY 1870 GGTGAGCAGCATGCGGAAGCCAGGCGCGCTGCTGACGTCCAGACTCCGCTTTCATCCC 1929
Db 3021 CAAGCTTGACGGCTGCGGCGGATTTGAACATGAGTGTGCTGGGAGCGCAGAACGTT 3080
QY 1930 CAAGCTTGACGGCTGCGGCGGATTTGAACATGAGTGTGCTGGGAGCGCAGAACGTT 1989
Db 3081 CCGCAGAGAAAAGAGGCGCGAGCTCTCAGCTCAGGGGTGAAGGCACTGTTTACGGTGT 3140
QY 1990 CCGCAGAGAAAAGAGGCGCGAGCTCTCAGCTCAGGGGTGAAGGCACTGTTTACGGTGT 2049
Db 3141 CAAGCTTGACGGCTGCGGCGGATTTGAACATGAGTGTGCTGGGAGCGCAGAACGTT 3200
QY 2050 CAAGCTTGACGGCTGCGGCGGATTTGAACATGAGTGTGCTGGGAGCGCAGAACGTT 2109
Db 3201 TATCCAGAGGCTGCGGCGGATTTGAACATGAGTGTGCTGGGAGCGCAGAACGTT 3260
QY 2110 TATCCAGAGGCTGCGGCGGATTTGAACATGAGTGTGCTGGGAGCGCAGAACGTT 2169
Db 3261 GCTGTACTTTTCAAGTGTGATGTACGGGCGGTACGACACCATCCCCCAGGACAGGCT 3320
QY 2170 GCTGTACTTTTCAAGTGTGATGTACGGGCGGTACGACACCATCCCCCAGGACAGGCT 2229
Db 3321 CAGGAGGCTATGCGCCAGCATCATCAAAACCCAGAACAGTGTGCTGGTTCGGTATGC 3380
QY 2230 CAGGAGGCTATGCGCCAGCATCATCAAAACCCAGAACAGTGTGCTGGTTCGGTATGC 2289
Db 3381 CGTGGTCCAGAGGCGCCCATGGGACGCTCCGCAAGGCTTCAAGAGCCACGCTCTTAC 3440
QY 2290 CGTGGTCCAGAGGCGCCCATGGGACGCTCCGCAAGGCTTCAAGAGCCACGCTCTTAC 2349
Db 3441 CTTGACAGACCTCCAGCGGTACATGCGGAGTGTGCTGGCTCACCTGAGGAGACAGCC 3500
QY 2350 CTTGACAGACCTCCAGCGGTACATGCGGAGTGTGCTGGCTCACCTGAGGAGACAGCC 2409

Db	3501	GCTGAGGGATGCCGTGCTCATGACGAGAGCTCTCCCTGAATGAGGCCACGACTGGCCT	3561
QY	2410	GCTGAGGGATGCCGTGCTCATGAGCAGAGCTCTCCCTGAATGAGGCCACGACTGGCCT	2469
Db	3561	CTTCCGAGCTCTCTACGCTTCATGTGCCACACGCCGTGCGCATCAGGGGCAAGTCCCTA	3620
QY	2470	CTTCCGAGCTCTCTACGCTTCATGTGCCACACGCCGTGCGCATCAGGGGCAAGTCCCTA	2529
Db	3621	CGTCCAGTGCACGGGATCCCGCAGGGCTCCATCCTCTCCACGCTGCTCTCAGCCCTGTG	3680
QY	2530	CGTCCAGTGCACGGGATCCCGCAGGGCTCCATCCTCTCCACGCTGCTCTCAGCCCTGTG	2589
Db	3681	CTACGGGGACATGGAGAACAGCTGTTGCGGGGATTCGGGGGAGCGGGCTGCTCTCGG	3740
QY	2590	CTACGGGGACATGGAGAACAGCTGTTGCGGGGATTCGGGGGAGCGGGCTGCTCTCGG	2649
Db	3741	TTTGGTGGATGATTTCTTGGTGACACCTCACCCTCACCCACGGGAAAACCTTCCTCAG	3800
QY	2650	TTTGGTGGATGATTTCTTGGTGACACCTCACCCTCACCCACGGGAAAACCTTCCTCAG	2709
Db	3801	GACCTTGGTCCGAGGTGCCCTGAGTATGGCTGCGTGTGAACTTGCGGAAGACAGTGGT	3860
QY	2710	GACCTTGGTCCGAGGTGCCCTGAGTATGGCTGCGTGTGAACTTGCGGAAGACAGTGGT	2769
Db	3861	GAACCTTCCCTGTAGAAAGACGAGGCCCTGGGTGGACGGCTTTTGTTCAGATSCGGGCCA	3920
QY	2770	GAACCTTCCCTGTAGAAAGACGAGGCCCTGGGTGGACGGCTTTTGTTCAGATSCGGGCCA	2829
Db	3921	CGGCCATTTCCTCGTGGGGCTGCTGCTGATACCCGGACCCCTGGAGGTGCAGACGA	3980
QY	2830	CGGCCATTTCCTCGTGGGGCTGCTGCTGATACCCGGACCCCTGGAGGTGCAGACGA	2889
Db	3981	CTACTCCAGTATGCCCGACCTCCATCAGAGCCAGTCTCACTTCAACCGCGGCTTCAA	4040
QY	2890	CTACTCCAGTATGCCCGACCTCCATCAGAGCCAGTCTCACTTCAACCGCGGCTTCAA	2949
Db	4041	GGCTGGGAGGAACATCGCTCGAAAACCTCTTTGGGGCTTTGGGGCTGAAGTGTCAAGCCT	4100
QY	2950	GGCTGGGAGGAACATCGCTCGCNAACCTCTTTGGGGCTTTGGGGCTGAAGTGTCAAGCCT	3009
Db	4101	GTTTCTGGATTTGCAGGTGAACAGCCTCCAGACGGTGTGCACCACACATCTACAAGATCCT	4160
QY	3010	GTTTCTGGATTTGCAGGTGAACAGCCTCCAGACGGTGTGCACCACACATCTACAAGATCCT	3069
Db	4161	CTGCTCTCAGGGGTACAGCTTTCACGCATGTGCTGCAGCTCCCATTTTCATCAGCAAGT	4220
QY	3070	CTGCTCTCAGGGGTACAGCTTTCACGCATGTGCTGCAGCTCCCATTTTCATCAGCAAGT	3129
Db	4221	TTGGAAGAACCCACATTTTTCTCGGGCTCATCTCTGACACGGCTCCCTCTGCTACTC	4280
QY	3130	TTGGAAGAACCCACATTTTTCTCGGGCTCATCTCTGACACGGCTCCCTCTGCTACTC	3189
Db	4281	CATCTCTGAAGCCAAAGACGAGGATGTGCTGGGGGCCAAGGGCGCGCGGCCCTCT	4340
QY	3190	CATCTCTGAAGCCAAAGACGAGGATGTGCTGGGGGCCAAGGGCGCGCGGCCCTCT	3249
Db	4341	GCCTTCGAGGGCGTGCATGCTGTCGCACCAACATTCCTGCTCAAGCTGACTCGACA	4400
QY	3250	GCCTTCGAGGGCGTGCATGCTGTCGCACCAACATTCCTGCTCAAGCTGACTCGACA	3309
Db	4401	CCGTGTCACTACGTGCCACTCCTGGGTCACCTCAGGACGCCCAGACGCGAGCTGACTCG	4460
QY	3310	CCGTGTCACTACGTGCCACTCCTGGGTCACCTCAGGACGCCCAGACGCGAGCTGACTCG	3369
Db	4461	GAAGCTCCGGGGACGACGCTGACTGCCCTGGAGGCCGACCAACCGGCACCTGCCCTC	4520
QY	3370	GAAGCTCCGGGGACGACGCTGACTGCCCTGGAGGCCGACCAACCGGCACCTGCCCTC	3429
Db	4521	AGACTTCAAGACCATCCTGGACT	4543
QY	3430	AGACTTCAAGACCATCCTGGACT	3452

[illegible]

QY 220 GGTGTGGTCCCTGGAGCAGCGGGCGCGCCCGCCCGCCCTCTCTTCGCGCAGGTGTC 279
Db 254 ATCCCTGAAGAGCTGTGTGCCAGAGGTGTGTCAGAGACTCTGCGAGGCGCAACGAGAGAAA 313
QY 280 CTGCTGAAGAGCTGTGTGCCAGAGGTGTGTCAGAGACTGTGCGAGCGCGCGCGAAGAA 339
Db 314 CGTGTGCTTTTGGCTTTGAGTGTCTTAACGAGGCCAGAGGGGCGCTCCCATGGCCTT 373
QY 340 CGTGTGCTTTGCTTGGCTTTCGCTGTGTGACGAGCGGGCGCGGGGGCGCCCGCAGGCGCT 399
Db 374 CACTAGTAGCGGTGCTAGCTACTGTGCCCACTGTATTGAGACCCCTGGTGTCACTGG 433
QY 400 CACCACAGGCTGGCAGCTACTGTGCCCAACACGGGTACCCAGCAGCTGCGGGGAGCGG 459
Db 434 TGCATGATGCTACTGTGTAGCGAGTGGCGGAGCAGCTGCTGTGTCTACCTGTGCGCACA 493
QY 460 GGGGTGGGGCTGCTGCTGCGCGCGGTGGGGGAGCAGCTGCTGTTCACCTGCTGGGACG 519
Db 494 CTGTGCTTTTATCTTCTGTGTGCCCCCAGCTGTGCTTACAGGTGTGTGGGTCTCCCT 553
QY 520 CTGCGGCTCTTTGTGTGTGCTCCAGCTGCGGCTTACAGGTGTGCGGGCGCGCGCT 579
Db 554 GTACCAAAATTTGCCACCAAGATATCTGGCCCTCTGTGCTCGCTAGTTACAGGCCAC 613
QY 580 GTACAGCTGCGGCTGCTCACTCAGGCGCGCGCGCCCGCCACAGCTAGTGGACCCGAG 639
Db 614 CCGACCGTGGGAGGAATTTCACTAACCTTAGTGTCTTAAACAGATCAAGAGCAGTAG 673
QY 640 GCGTCTGGGATGCGACGGGCTTGGAACTATAGGTC--AGGG-AGCGCGGGTCCCGCT 696
Db 674 TCGCCAGGAAGACCCGAAACCCCTGGCCCTTGCCTCTCGAGGTACAAGAGCAGCTGAG 733
QY 697 GGGCTTGCACAGCCCGGGTGGAGGAGCGGGGGGAGTGCACAGCAAGCTGTGCGCTT 756
Db 734 TCTCACCAGTACAGTGTGCTTCTAGTATAGAGGCCAGTATCTCTGTCGAGAGT 793
QY 757 GGCNAAGAGCCAGCGCTGGCGCTGCGCCGTGAGCGGAGCGGCGCGCTTGGGC-AGG 815
Db 794 GGAGGAGGACCCACAGGAGGTGTCTACCAACCCATCAG-GCAATCATGGTGCACAA 852
QY 816 GGTCTTGGGCGCACCGGGCAGACGCGGTGGACCGAGTACCGTGGTCTGTGTGTGT 875
Db 853 GTCTCTGCTCCGTCGCCGAGGTGCCTACTCGAGAGAAAGATTTCTCTTAAAGGAAAG 912
QY 876 CACCTGCCAGACCCCGGAAGAACCCCTCTTTGGAGGTGCG-CT-CTCT-GGCACG 932
Db 913 TGTCTACCTGAGTCTCTGTGGTGTGCTGTGTAAACAGACCCAGCTCCACATCTC 972
QY 933 -GCCACTCCACCCATCCGTGTGGCGGCA-GCACACGCGGGCGCCCA-TCCACATCGC 989
Db 973 TGTGTCAACACCCCGCAAAATGCTTTTTCAGCTCAGGCCATTTATTGAGACCAAGATT 1032
QY 990 GGGCACACCTCCCTGGGACAC-GCCTTGTGCCCC-GGTG-TAGCGCGAGCAAGCACT 1046
Db 1033 TCCTTTACTCCAGGGAGATGGCAAGAGGTCTTAAACCCCTCATCTCTACTCAGCAACC 1092
QY 1047 TCCTCTACTCTCAGGGA---CAAGGAGCAGCTGCGGCGCTCTCTCTACTCAGCTCTC 1103
Db 1093 TCCAGCCTTAATTTGACGTGGGCGCAGGAGACTGTGTGAGATCATCTTTCTGGGCTCAAGGC 1152
QY 1104 TGAGGCGCCAGCCTGACTGTGGCTCGGAGGCTGTGTGAGACCATCTTTCTGGGTTCCAGGC 1163
Db 1153 CTAGGACATCAGACCACTCTGAGGACACACCGTCTATCGCTCGATACTGGCAGATGC 1212
QY 1164 CTTGGATGCCAGGAGCTCCCGCAGGTGTGCGCGCTTGCAGGCTGCTGCGCAATGC 1223
Db 1213 GGGCCTGTTCACACAGCTGTGTGTAACCATGAGAGTGGCAATA-----TG-----TCA 1263
QY 1224 GGGCCTGTTCCTGAGCTGTTGGGAACCAACAGCGGAGTGGCCCTACGGGTGCTCTCTCA 1293
Db 1264 ----G-ACT--C-CT-C-AG-----GTCACATT-GCAG---GTTTC---GAACAGCA-A 1299
QY 1284 AGAGCAGCTGCCGCTGCGAGCTGCGGTACCCCGCAGCAGCGGTGTGTGTCGCCGGGAGA 1343

Db 1300 ACCAACAGG-----TGACAGATGCCCTTG-A--AC-AC-CAGCCCCACCG-CACCTCATGG 1347
QY 1344 AGCCCCAGGGCTCTGTGGCGCCCCCGAGGAGGAGACACAGACCCCGCTGCGCTGGTGC 1403
Db 1348 ATTGTCTCCGGCTGCACAGAGTCCCTGGCAGGTATATGGTTTCTTCGGGCGCTGTCTCT 1407
QY 1404 AGCTGCTCCGCGACAGCAGCAGCCCCCTGGCAGGTGTACGGCTTCTGCGGGCGCTGCTGC 1463
Db 1408 GCAAGTGGTGTCTGTCTGTGGGTACAGGSCAAATGAGCGCGCTCTTTAAGA 1467
QY 1464 GCGGCTGTGCCCCCAGGCTCTGGGGCTCCAGGCAACGAAAGCGGCTCTCTCAGGA 1523
Db 1468 ACTTAAAGAAAGTTTCATCTCGTTGGGAAATACGGCAAGCTATCACTGCAGGAATCTGT 1527
QY 1524 ACACCAAGAGTTTCATCTCCCTGGGAGCATGCCAAGCTCTCGCTGCAGGAGCTGACGT 1583
Db 1528 GGAAGATGAAGTAGAGATTGGCCACTGGCTCCGAGAGAGCCCGGGGAAGACCGTGTCC 1587
QY 1584 GGAAGATGAGCGTGGGAGCTGCGCTTGGCTGCGCAGGAGCCAGGGTTGGCTGTGTTTC 1643
Db 1588 CCGCTGCAGAGCACCGTCTGAGGAGAGGATCCTGGCTACCTTCTGTCTGGCTGATGG 1647
QY 1644 CGGCGGAGAGACCGTCTGCTGAGGAGATCCTGGGCAAGTTCTCGACTGGGCTGATGA 1703
Db 1648 ACATACCTGCTGCTAGCTGTAGGTCATCTTTTACATCACAGAGAGCATTTCCAGA 1707
QY 1704 GTGTGTAGCTGCTGAGCTGCTCAGGCTCTTCTTTATGTACAGGAGACCGTTTCAA 1763
Db 1708 AGAACAGGCTCTTCTTCTACCGTAAGAGTGTGTGAGCAAGCTGCAGAGCATTTGGAGTCA 1767
QY 1764 AGAACAGGCTCTTCTTCTACCGTAAGAGTGTGTGAGCAAGTTTCAAAGCATTTGAAATCA 1823
Db 1768 GGCACACCTTGAGAGAGTGGGCTACGGGAGCTGTACAGAGAGAGTCTAGGCATCAC 1827
QY 1824 GACAGCACTTGAAGAGGTGCAGCTGGGGAGCTGTGGAAAGCAGAGGTTCAGGAGCATC 1883
Db 1828 AGGACACTGTGCTAGCTAGCCATGCCATCTGCAGACTGCGCTTCATCCCCAAGCCCAACGCC 1887
QY 1884 GGGAGCCAGGCGCGCTGCTGAGCTCCAGACTCCGGCTTCATCCCCAAGCTTCAGCGGC 1943
Db 1888 TCGCGCCCATTTGAACATAGTTATAGCATGGGTACAGAGCTTTGGGCAAGAAAGC 1947
QY 1944 TCGCGCGATTTGAACATGGACTACGTGCTGGGAGCCAGAACGTTCCGCAGAGAAAAGA 2003
Db 1948 AGGCGCAGCATTTACCCAGCGCTCTCAAGACTCTTTCAGCATGCTCAACTATGAGCGGA 2007
QY 2004 GGGCGGAGCTCTCACCTCGAGGGTGAAGGCACTGTTTCAGCGTGTCTCAACTACGAGCGGG 2063
Db 2008 CAAACATCTCTACCTTATGGGTCTTCTGTACTGGGTATGAATGACATCTACAGGACCT 2067
QY 2064 CGCGCGCGCGCGCTCTCTGGGCGCTCTGTGCTGGGCTGGACGATATCCACAGGGCT 2123
Db 2068 GCGCGCGCTTTGTGCTGCTGTGCTGTCTGGACACACACCCAGGATGTACTTTGTTA 2127
QY 2124 GGCACCTTCTGCTGCTGTGCGGGCCAGAGACCCCGCGCTGAGCTGTACTTTGTCA 2183
Db 2128 AGGAGATGTACCGGCGCTTATGATCCATCCCGCAGGGTAAAGCTGTGTGAGGTGTGTT 2187
QY 2184 AGTGTGATGTACGGGGCGGTACGACACCATCCCGCAGGACGGCTCACGGAGGTCTACG 2243
Db 2188 CCAATATGATCAGGCACTCGGAGAGCAGTACTGTATCCCGCAGTATGACATGTGTCGCGGA 2247
QY 2244 CCAGCATCATAAACCC--CAGA-A-CAGGTACTGCGTGGGTATGCCGTGGTGGTCCAGA 2300
Db 2248 GAGATAGCCAGGCAAGTCCACAAAGTCTTTAGGAGACAGGTACACACCTCTCTGACC 2307
QY 2301 AGCGCGCCCATGGCAGCTCCGCAAGGCTTCAAGAGCCAGCTCTCTACCTTGCAGACC 2360
Db 2308 TCCAGCCATACATGGGCGAGTTCTTAAAGCATCTGCAGATTTCAGATGCCAGTGCACCTGA 2367
QY 2361 TCCAGCGGTACATGCGACAGTTCGTGGCTCACCTGCAGGA-----GA---CCAGCGCGGTGA 2414

[illegible]

QY	2689	CCACGCGAAAACCTTCTCCTCAGGACCTGGTCCGAGGTGTCCCTGAGTATGGCTCGCTGGT	2748
Db	126	AMTSRNRTGKTANNVDSRNMGDASVGS-D-KWTKKHANSADGKVGSKNNGDRNRVGTG	184
QY	2749	GAACCTTGGCGAAGACAGTAGTGGTGAACATTCCTCTGTAAGAAGCAGGCGCCTGGGTGGCACGGC	2808
Db	185	TKSNVSNCGGKNKRDVSSYANNKCC	210
QY	2809	TTTGTTCAGATCCGCGCCACACGGCC	2834
RESULT	9		
LOCUS	128278	215 bp	DNA
DEFINITION	Sequence 5 from patent US 5569830.		PAT
ACCESSION	128278		
NID	g1819054		
KEYWORDS	Unknown.		
SOURCE	Unknown.		
ORGANISM	Unknown.		
REFERENCE	1 (bases 1 to 215)		
AUTHORS	Bennett,A., Labavitch,J.M., Powell,A. and Stotz,H.		
TITLE	Plant inhibitors of fungal polygalacturonases and their use to control fungal disease		
JOURNAL	Patent: US 5569830-A 5 29-OCT-1996;		
FEATURES	Location/Qualifiers		
source	1..215		
BASE COUNT	15 a	8 c	25 g
ORIGIN	/organism="unknown"	26 t	141 others
Query Match	0.9%	Score 37;	DB 17; Length 215;
Best Local Similarity	14.2%	Pred. No. 5.00e-06;	
Matches	22; Conservative	64; Mismatches	68; Indels 1; Gaps
Db	50	YRVNDSGHKNKYSANNVGGNNVGNNAKTHYYTHTVNSGADSKVTYDSYNASGCTSSSNGG	109
Cp	1128	CGAGCCCGCATTCAGGCTGGCGCTTCAGAGAGCTGAGTAGGAAGAGGCGCCGACGCTGCTCC	1069

Db	110	TDGNSRGADSTGSSKTA-WTSENRTGKTANNANVDSRNNMGDASVSGDKNTKHKAKNSADGK	168
l:	:	:	:
Cp	1068	TTGTGCGCTGAGGAGTAGAGGAAGTGTGTGCTCGCGGTACACCGGGGACAAAGCGCTG	1009
l:	:	:	:
Db	169	VGSNNNGDRNNRYGTGKTSNYSNNCGGCKRDRYSS	203
l:	:	:	:
Cp	1008	TCCAGGACGTGTGGCCGGAGTGGATGGGG	974
l:	:	:	:
RESULT	10		
LOCUS	E04076	565 bp	RNA
DEFINITION	gDNA encoding envelope region of type C hepatitis virus.		
ACCESSION	E04076		
NID	92172286		
KEYWORDS	JP 1992349885-A/1.		
SOURCE	Hepatitis C virus.		
ORGANISM	Hepatitis C virus		
REFERENCE	Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae; Hepatitis C-like viruses.		
AUTHORS	1 (bases 1 to 565)		
TITLE	TSutae,M., Kazuaki,C., Hiromitsu,K. and Yataro,I..		
JOURNAL	NUCLEIC ACID FRAGMENT WITHIN ENVELOPE REGION OF HEPATITIS C VIRUS AND METHOD FOR DETECTING THE SAME		
COMMENT	Patent: JP 1992349885-A 1 04-DEC-1992;		
TEIJIN LTD			
OS	Hepatitis C virus		
PN	JP 1992349885-A/1		
PD	04-DEC-1992		
PF	29-MAY-1991	JP 1991152169	
PI	MORINAGA TSUTAE, CHAYAMA KAZUAKI, KUNADA HIROMITSU, PI		
ICHIKAWA YATARO			
PC	C12H15/10.C12Q1/68.C12Q1/70//C12N15/11;		
CC	strandedness: Single;		
CC	topology: Linear;		
FEATURES	Location/Qualifiers		
source	1..565		
BASE COUNT	60 a	93 c	107 g
ORIGIN	/organism="Hepatitis C virus" /db_xref="taxon:11103"		
Query Match	0.9%;	Score 36;	DB 17; Length 565;
Best Local Similarity	31.7%;	Pred. No. 1.98e-05;	
Matches	59;	Conservative	58; Mismatches 66; Indels 3; Gaps 3;
Db	233	GYYRBCYGTGTGCGSARRRYNNYTCBMGNTGYTGGYRGCCTBACYCCCAGCS	292
l:	:	:	:
1561	GCTTCGCTGAGGAGCTGAGAGTGGCGTGGCGGACTCGCTTGGCTGGCGAG	1620	
l:	:	:	:
Db	293	TYGVRGVAGRRVBYAVYVYCCRCBRYGVNRMTWCGBCGBCAYATCGAYTGTCTG	352
l:	:	:	:
Qy	1621	GAGCC-CAGGGGTGGCTGTCTTCGGCGCGAGACACCGTCTCGGTGAGGAGATCCTGG	1679
l:	:	:	:
Db	353	THGGR-SVGCYRYTCTGTCTGCGYVTSYATGKRGAGYATBTGCGGRCTCYGTYYT	411
l:	:	:	:
Qy	1680	CCAAATTCTGCACCTGGCTGATGAG-TGTGTACGTGCTGCGAGCTCTCAGGTCTTCTTT	1738
l:	:	:	:
Db	412	CTYRTY	417
l:	:	:	:
Qy	1739	TATGTC	1744
l:	:	:	:
RESULT	11		
LOCUS	A10161	201 bp	DNA
DEFINITION	Synthetic DNA for preprotease leader & prochymosin.		
ACCESSION	A10161		
NID	9490655		
KEYWORDS	KEYWORDS		
SOURCE	unidentified.		
ORGANISM	unidentified.		
REFERENCE	1 (bases 1 to 201)		

Q51746;
AC 31-MAY-1994 (first entry)
DE Oligonucleotide probe MK14-A
KW Oligonucleotide; DNA probe; mycobacteria; disease diagnosis;
KW ss.
OS Synthetic.
PN EP-571911-A.
PD 01-DEC-1993.
PF 24-MAY-1993; 108325.
PR 26-MAY-1992; US-889651.
PA (BECT) BECTON DICKINSON CO.
PI Shank DD, Spears PA;
DR WPI: 93-378844/48.
PT New oligo:nucleotide probes specific for Mycobacteria - used for
PT detection and amplification of Mycobacteria nucleic acid in
PT samples
PS Claim 3; Page 14; 23pp; English.
CC Oligonucleotide probe MK14-A consists of nucleotides 5-95 of MK14
CC (Q51735). It hybridized to all spp. of mycobacteria tested, but
CC cross reacted to a few non-mycobacterial spp. The probe may
CC be useful as an initial screen for mycobacterial infection.
CC See also Q51735-45 and Q51747-59.
SQ Sequence 91 BP; 5 A; 17 C; 15 G; 4 T;

Query Match 1.1%; Score 45; DB 9; Length 91;
Best Local Similarity 11.9%; Pred. No. 2.21e-10;
Matches 7; Conservative 45; Mismatches 7; Indels 0; Gaps 0;

Db 4 tccggcgssvhsyvvhvshhhsvvvhhvhvsvvvhhvhvhhvhyvsvct 62
||| ||| :||||:||||:||||:||||:||||:||||:||||:
Cp 1118 TCAGCGTGGGCTCAGAGAGCTGAGTAGGAGGAGGCCGCAGCTGCTTCGTGCCT 1060

RESULT 3
ID TT6405 standard; DNA; 178 BP.
AC T76405;
AD T76405;
DT 15-SEP-1997 (first entry)
DE Human endothelin-1 antisense oligonucleotide.
KE Asthma; airway epithelium; adenosine free; cystic fibrosis;
KW chronic obstructive pulmonary disease; bronchitis; ss.
OS Synthetic.
PN W09640162-A1.
PD 19-DEC-1996.
PF 06-JUN-1996; U09306.
PR 07-JUN-1995; US-474497.
PA (UYEC-) UNIV EAST CAROLINA.
PI Metzger WJ, NYCE JW;
WPI: 97-051871/O5.
PT Treatment of airway diseases such as asthma - by topically applying
PT adenosine-free antisense oligo:nucleotide to airway epithelium of
PT subject
PS Claim 5; Page 38; 71pp; English.
CC A method for treating airway disease in a subject has been produced,
CC which involves the topical administration of an essentially adenosine
CC free antisense oligonucleotide (ON) to the airway epithelium of the
CC subject. The present sequence is an antisense oligonucleotide specific
CC for the human endothelin-1, targeted at the initiation codon. The
CC method can be used to treat airway diseases such as cystic fibrosis,
CC asthma, chronic obstructive pulmonary disease, bronchitis and other
CC airway diseases characterised by an inflammatory response. By
CC eliminating adenosine from the antisense ON, its liberation upon
CC antisense degradation is prevented, thereby preventing adenosine-
CC induced bronchoconstriction in patients with hyper-reactive airways.
SQ Sequence 178 BP; 0 A; 52 C; 46 G; 32 T;

Query Match 1.1%; Score 45; DB 32; Length 178;
Best Local Similarity 37.3%; Pred. No. 2.21e-10;
Matches 44; Conservative 40; Mismatches 33; Indels 1; Gaps 1;

Db 53 bccbcgcgactbgctcbgcgctbbgbctgtgttctgtttcttggbgtccttcgcbgcgc 112
QY :||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:
44 GCCACC CCCCGGAGTGGCGCGCTGCCCGTGCCGAGCCGTGCGCTGCTTCGTGCAGC 103


```

Db      73  ccbbygbcbbtbtccgcbggctgtgcbgbbcbtcbtgggttcgbbggtgtbgtggb 132
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy      92  CTGTCGGCAGCCACTACCGCGGAGTGCTCCGCTGCCACGATTC-GTGGCGGCGCTGG 150
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      133  qbgtttggggbbgggtctgbgtccbccgggbbgbbgtbtcbtcbtggbbgcbtbgcgg 192
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy      151  GCCCAGCGCTGCGGGTGTGTGACGCGGGGACCGCGCGGCTTCCTCCGCGCTGGTGGC 210
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      193  tbbgcbctbctgtctgtcbcbcbccccctctgcbgcb 231
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy      211  CCAGTGCTGTGTGTGCTGCCCTGGGACGACGCGCGCC 249

RESULT 12
ID      Q70465 standard; DNA; 114 BP.
AC      Q70465;
DT      05-APR-1995 (first entry)
DE      Generic DNA sequence to generate a random TSAR peptide library.
KW      TSAR: totally synthetic affinity reagent; synthetic; binding domain;
KW      effector domain; concatenated heterofunctional protein; linker;
KW      direct; rapid; detection; screening; treatment; generic; ss.
OS      Synthetic.
FH      Key
FT      Location/Qualifiers
FT      misc_feature 55..60
FT      /*tag= a
FT      /note= "this sequence represents '2'; 2 can be a
FT      sequence of 6, 9 or 12 nucleotides (see
FT      comments)"
FT      FT
PN      WO9418318-A.
PD      18-AUG-1994.
PF      01-FEB-1994.
PP      000977.

```

30-DEC-1999; US-176300.
PR 31-JAN-1994; US-189331.
PA (UYNC-) UNIV NORTH CAROLINA.
PI Fowlkes DM, Kay BK;
PT WPI: 94-279739/34.
DR P-PSDB: R65150 and R65151.
PT Identifying proteins or peptide(s) which bind a ligand - by
PT screening a recombinant vector library expressing fusion proteins
PT comprising a binding domain and an effector domain
PS Disclosure; Page 35; 255pp; English.
CC Q70465 is a generic DNA sequence used to generate random TSAR (Totally
CC Synthetic Affinity Reagents) peptides.This generic formula can also be
CC represented as follows: X(NNB)5(TGC)(NNB)11Z(NNB)14(TGC)(NNB)3Y.X
CC and Y are flanking restriction sites (X is not the same as Y) that are
CC not specified further. Other generic sequences are shown in Q70466-68.
CC Other specific peptides generated by these generic sequences are shown in
CC R65151-54. TSARs are concatenated heterofunctional proteins or peptides,
CC comprising at least two functional regions - a binding domain with
CC affinity for a ligand and a second effector peptide portion that is
CC chemically or biologically active.They may further comprise a linker
CC peptide between the 2 domains.The oligonucleotides are also designed so
CC that the expressed peptide contains 2 or 4 cysteine residues positioned
CC in, or flanking, the unpredicted or variant residues. These residues
CC confer some degree of conformational rigidity to the peptides. The TSARs
CC or compns. comprising a TSAR binding domain can be used in vivo to
CC deliver a chemically or biologically active moiety, eg. metal ion,
CC radioisotope, peptide, toxin or enzyme, to the specific target or on the
CC cell. They can also replace the function of macromolecules, eg.
CC monoclonal or polyclonal antibodies and therefore circumvent the need
CC for complex methods of hybridoma formation or in vivo antibody
CC production. The TSARs are easily characterised and have designed
CC activity allowing direct and rapid detection in a screening process.

```

SQ sequence 114 BP; 0 A; 2 C; 2 G; 2 T;
Query Match 0.98; Score 38; DB 12; Length 114;
Best Local Similarity 5.4%; Pred. No. 1.23e-06;
Matches 6; Conservative 33; Mismatches 73; Indels 0; Gaps 0;
Db 3 bnnbnnbnnbnnbgnbnnbnnbnnbnnbnnbnnbnnbnnbnnnnnn 62
: : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

Matches 6; Conservative 33; Mismatches 73; Indels 0; Gaps 3

[illegible]

Search completed: Mon Aug 3 16:47:21 1998
Job time : 1666 secs.

(TM)

Release 3.1A John F. Collins, Blocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_nn n.a. - n.a. database search, using Smith-Waterman algorithm
Run on: Mon Aug 3 10:56:05 1998; MasPar time 4562.35 Seconds
1232.517 Million cell updates/sec
Bular output not generated.

Title: >US-08-912-951-1
Description: (1-4015) from US08912951.seq
Perfect Score: 4015
N.A. Sequence: 1 GCAGCGCTCGCTCCTGCTGC.....TTTTTCAGTTTGAAGAAAAA 4015
Comp: CGTCGGAGCAGGACGACG.....AAAAAGTCAAACTTTTTT

Scoring table: TABLE default
Gap 6

Nmatch STD : Dbase 0; Query 0

Searched: 1832099 seqs, 700269816 bases x 2

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: emb1-est55
Database: genbank-est107

3:gb_est1 4:gb_est10 5:gb_est11 6:gb_est12 7:gb_est13
8:gb_est14 9:gb_est15 10:gb_est16 11:gb_est17
12:gb_est18 13:gb_est19 14:gb_est20 15:gb_est21
16:gb_est22 17:gb_est23 18:gb_est24 19:gb_est25
20:gb_est26 21:gb_est27 22:gb_est28 23:gb_est29 24:gb_est30
25:gb_est31 26:gb_est32

Statistics: Mean 12.222; Variance 2.606; scale 4.689

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	385	9.6	389	8	AA281296	zt08g02.r1 NCI_CGAP_GC
2	315	7.8	340	15	AA811084	oe85c05.s1 NCI_CGAP_GC
3	285	7.1	315	15	AA748707	ny02e05.s1 NCI_CGAP_GC
4	248	6.2	409	4	AA311750	EST182469 Jurkat T-cel
5	67	1.7	247	13	AA754458	97SN1784 Rice Immature
6	65	1.6	252	13	AA754459	97SN1787 Rice Immature
7	62	1.5	247	13	AA754458	97SN1784 Rice Immature
8	51	1.3	247	13	AA754458	97SN1784 Rice Immature
9	52	1.3	303	4	AA299878	EST12462 Uterus tumor
10	33	0.8	375	25	AA200728	mul3h09.r1 Soares 2NDM
11	31	0.8	660	12	AF034177	Homo sapiens ntcon6 co
12	33	0.8	2275	12	AF034173	Homo sapiens ntcon2 co
13	33	0.8	2275	12	AF034173	Homo sapiens ntcon2 co

14	27	0.7	317	26	AQ008007	CIT-HSP-2288C3.TF CIT-	6.67e-06
c	15	0.7	560	13	AA075218	ac79e03.s1 Stratagene	6.67e-06
	16	0.7	660	12	AF034177	Homo sapiens ntcon6 co	6.67e-06
	17	0.7	1287	13	AF038250	Homo sapiens clone ntc	4.63e-07
c	18	0.7	1287	13	AF038250	Homo sapiens clone ntc	4.63e-07
	19	0.6	148	10	AA619013	vo66g06.r1 Soares mous	1.25e-02
	20	0.6	181	5	AA386387	EST81369 Prostate gnan	1.10e-03
	21	0.6	187	15	AA855630	vw68h11.r1 Stratagene	1.10e-03
	22	0.6	213	14	R72797	Y109c09.r1 Homo sapien	1.10e-03
c	23	0.6	275	16	AA899633	UI-R-80-dg-f-09-0-UI.s	1.25e-02
	24	0.6	278	15	AA882818	TENS0393 T. cruzi epim	1.10e-03
c	25	0.6	288	7	C27729	Rice cDNA, partial seq	1.25e-02
c	26	0.6	288	24	AA077661	7B35F02 Chromosome 7 F	1.25e-02
	27	0.6	301	17	AA848961	EST191723 Normalized r	1.10e-03
	28	0.6	318	19	H50134	yo27a07.r1 Homo sapien	1.10e-03
	29	0.6	318	17	AA799774	EST189271 Normalized r	1.10e-03
	30	0.6	343	23	AA098755	T3985 MVAT4 bloodstrea	8.93e-05
	31	0.6	359	4	AA030595	EST16305 Aorta endothe	1.10e-03
	32	0.6	382	14	R54656	YJ74d04.r1 Homo sapien	1.10e-03
	33	0.6	396	20	T26788	T529 Trypanosoma bruce	8.93e-05
	34	0.6	400	19	H44707	YP24c08.r1 Homo sapien	8.93e-05
c	35	0.6	429	14	R55367	YG88h01.s1 Homo sapien	1.25e-02
c	36	0.6	432	5	AA043978	ZK38f01.s1 Soares preg	1.10e-03
	37	0.6	438	17	AA924206	UI-R-A1-dx-a-11-0-UI.s	1.10e-03
	38	0.6	463	26	AQ000405	CIT-HSP-2288L6.TR CIT-	1.10e-03
c	39	0.6	478	17	AA925372	UI-R-A1-ee-d-07-0-UI.s	1.10e-03
	40	0.6	487	19	R86860	YM86a03.r1 Homo sapien	1.10e-03
	41	0.6	496	10	AA429394	ZW32h12.s1 Soares ovar	1.25e-02
c	42	0.6	511	18	AA997735	UI-R-CO-hn-b-08-0-UI.s	1.10e-03
c	43	0.6	540	16	AA901048	UI-R-A1-dv-e-09-0-UI.s	1.10e-03
	44	0.6	590	10	AA622227	no43c05.s1 NCI_CGAP_Pr	1.10e-03
c	45	0.6	705	17	AA941159	LD25143.5prime LD Dros	1.10e-03

ALIGNMENTS

RESULT	1	AA281296	389 bp	mrna	EST	14-AUG-1997
LOCUS		zt08g02.r1 NCI_CGAP_GCB1	Homo sapiens	cdna	clone IMAGE:712562	5'
DEFINITION		mrna sequence.				
ACCESSION		AA281296				
NID		91924194				
KEYWORDS		EST.				
SOURCE		human.				
ORGANISM		Homo sapiens				
REFERENCE		1 (bases 1 to 389)				
AUTHORS		NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.				
TITLE		National Cancer Institute, Cancer Genome Anatomy Project (CGAP),				
JOURNAL		Tumor Gene Index				
COMMENT		Unpublished (1997)				
		Contact: Robert Strausberg, Ph.D.				
		Tel: (301) 496-1550				
		Email: Robert.Strausberg@nih.gov				
		This clone is available royalty-free through LLNL; contact the				
		IMAGE Consortium (info@image.llnl.gov) for further information.				
		Insert Length: 2187 Std Error: 0.00				
		Seq primer: -28m13 rev2 ET from Amersham				
		High quality sequence stop: 385.				
FEATURES		Location/Qualifiers				
source		1. 389				
		/organism="Homo sapiens"				
		/note="Vector: pT73D-Pac (Pharmacia) with a modified				
		polylinker: Site_1: Not I; Site_2: Eco RI; 1st strand cDNA				
		germline center B cells by flow sorting (CD20+, IgD-),				
		provided by Dr. Louis M. Staudt (NCI), Dr. David Allman				
		(NCI) and Dr. Gerald Marti (CBER). cDNA synthesis was				
		primed with a Not I - oligo(dT) primer				
		[5'-TGTTACCAATCTGAGTGGAGCGCGCTCATTTTTTTTTTTTTT-				

3'). Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library was constructed by Berto Soares and M. Fatima Bonaldo.

/db_xref="taxon:9606"
/clone="IMAGE:712562"
/clone_lib="NCI_CGAP_GCB1"
/tissue_type="germinal center B cell"
/lab_host="DH10B"

<1. ->389

87 a 102 c 123 g 77 t

Query Match 9.6%; Score 385; DB 8; Length 389;

Best Local Similarity 99.5%; Pred. No. 0.00e+00;

Matches 387; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

1 GCCAAGTTCCTGCACTGGCTGATGAGTGTGTACGTCGTCGAGCTGCTCAGGTCTTCTTT 60

1679 GCCAAGTTCCTGCACTGGCTGATGAGTGTGTACGTCGTCGAGCTGCTCAGGTCTTCTTT 1738

61 TATGTACAGGAGACACGCTTTCACAAAGACAGGCTCTTTTCTACCGAAGAGTGTCTGG 120

1739 TATGTACAGGAGACACGCTTTCACAAAGACAGGCTCTTTTCTACCGAAGAGTGTCTGG 1798

121 AGCAAGTTGCAAGCAATGGAATCAGACAGCATTGAAGGGTGCAGCTGCGGGACGTG 180

1799 AGCAAGTTGCAAGCAATGGAATCAGACAGCATTGAAGGGTGCAGCTGCGGGACGTG 1858

181 TCGGAGCAGAGTCAGGACGATCGGGAGCCAGGCGCCCTGCTGACGTCCAGACTC 240

1859 TCGGAGCAGAGTCAGGACGATCGGGAGCCAGGCGCCCTGCTGACGTCCAGACTC 1918

241 CGCTTCATCCCAAGCTGACGGCTCGCGCGATTGTGAACATGGACTAGCTCGTGGGA 300

1919 CGCTTCATCCCAAGCTGACGGCTCGCGCGATTGTGAACATGGACTAGCTCGTGGGA 1978

301 GCCAGAACGTTCCGACAGAGAAAGAGGCGCGAGCTCTCACCCTCGAGGGTGAAGGCAC 360

1979 GCCAGAACGTTCCGACAGAGAAAGAGGCGCGAGCTCTCACCCTCGAGGGTGAAGGCAC 2038

361 TTCAGCGTCTCAACTACGAGCGGCGCG 389

2039 TTCAGCGTCTCAACTACGAGCGGCGCG 2067

RESULT 2
LOCUS AA811084 340 bp mRNA EST 19-FEB-1998
DEFINITION oa85c05.s1 NCI_CGAP_GCB1 Homo sapiens cDNA clone IMAGE:1319048,
mRNA sequence.

ACCESSION AA811084

NID 92880695

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;

Primates; Catarrhini; Homidae; Homo.

1 (bases 1 to 340)

NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

Tumor Gene Index

Unpublished (1997)

CONTACT: Robert Strausberg, Ph.D.

Tel: (301) 496-1550

Email: Robert.Strausberg@nih.gov

Tissue Procurement: Louis M. Staudt, M.D., David Allman,

Ph.D., Gerald Marti, M.D.

cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima

Bonaldo, Ph.D.

cDNA Library Arrayed by: Greg Lennon, Ph.D.

DNA Sequencing by: Washington University Genome Sequencing Center

Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.llnl.gov/brp/image/image.html

Insert Length: 2249 Std Error: 0.00

Seq primer: -40ml3 fwd. ET from Amersham

High quality sequence stop: 331.

Location/Qualifiers

1. .340

/organism="Homo sapiens"

/note="Vector: pT73D-Pac (Pharmacia) with a modified

polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA

was prepared from human tonsillar cells enriched for

germinal center B cells by flow sorting (CD20+, IgD+),

provided by Dr. Louis M. Staudt (NCI), Dr. David Allman

(NCI) and Dr. Gerald Marti (CBER). cDNA synthesis was

primed with a Not I - oligo(dT) primer

[5'-TCTTACCAATCTGAAGTGGAGGCGGCTCATTTTCTTTTCTTTT-

3']. Double-stranded cDNA was ligated to Eco RI adaptors

(Pharmacia), digested with Not I and cloned into the Not I

and Eco RI sites of the modified pT73 vector. Library

went through one round of normalization, and was

constructed by Bento Soares and M. Fatima Bonaldo."

/db_xref="taxon:9606"

/clone="IMAGE:1319048"

/clone_lib="NCI_CGAP_GCB1"

/tissue_type="germinal center B cell"

/lab_host="DH10B"

BASE COUNT 73 a 86 c 100 g 81 t

ORIGIN

Query Match 7.8%; Score 315; DB 15; Length 340;

Best Local Similarity 99.1%; Pred. No. 0.00e+00;

Matches 328; Conservative 0; Mismatches 1; Indels 2; Gaps 2;

Db 12 TTTTTCCTCAAACTGAAACTCATATATTCAGTATTTTACTCCACAGACCTCCCCC 71

Cp 4015 TTTTTCCTCAAACTGAAACTCATATATTCAGTATTTTACTCCACAGACCTCCCCC 3956

Db 72 AATTGACCCACAGG-ACCCCATCCAGGTGCGGCTCTGTACAGGCGACAC 130

Cp 3955 AATTGACCCACAGGACCCCATCCAGGTGCGGCTCTGTACAGGCGACAC 3896

Db 131 CTTTGTGCTACTCCAAATCCAGAGTCCAGAGGTCCTTCTCAGGGTCTCCACCTGGATG 190

Cp 3895 CTTTGTGCTACTCCAAATCCAGAGTCCAGAGGTCCTTCTCAGGGTCTCCACCTGGATG 3836

Db 191 GTTGGGTGGAAGCAAGGAGGCGAGGCGGAGGCGGAGTGAACATGGGGAATCTGGGATG 250

Cp 3835 GTTGGGTGGAAGCAAGGAGGCGAGGCGGAGGCGGAGTGAACATGGGGAATCTGGGATG 3776

Db 251 GACTATTCTCTATGTGGGAGTGGAAAGCCGGCTCTCTGTGAGGAAAGAGTGGCCCTGGGG 310

Cp 3775 GACTATTCTCTATGTGGGAGTGGAAAGCCGGCTCTCTGTGAGGAAAGAGTGGCCCTGGGG 3716

Db 311 TGGAGCGGAGCGCA-CCTGTGGGGAAGTGA 340

Cp 3715 TGGAGCGGAGCGGAGGCGGAGGCGGAGGCGGAGTGAACATGGGGAATCTGGGATG 3685

RESULT 3

LOCUS AA748707 315 bp mRNA EST 18-FEB-1998

DEFINITION oy02e05.s1 NCI_CGAP_GCB1 Homo sapiens cDNA clone IMAGE:1270592,

mRNA sequence.

ACCESSION AA748707

NID 92788665

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;

Primates; Catarrhini; Homidae; Homo.

1 (bases 1 to 315)

NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
JOURNAL Tumor Gene Index
COMMENT Unpublished (1997)

Contact: Robert Strausberg, Ph.D.
 Tel: (301) 496-1550
 Email: Robert.Strausberg@nih.gov
 Tissue Procurement: Louis M. Staudt, M.D., Ph.D., David Allman,
 Ph.D., Gerald Marti, M.D.
 cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
 Bonaldo, Ph.D.
 cDNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 www-bio.llnl.gov/bbrp/image/image.html

Insert Length: 1226 Std Error: 0.00
 Seq primer: -40ml3 fwd. Et from Amersham
 High quality sequence stop: 281.

FEATURES
 Location/Qualifiers
 1..315
 /organism="Homo sapiens"
 /note="Vector: pT7R3D-Pac (Pharmacia) with a modified
 polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
 was prepared from human tonsillar cells enriched for
 germinal center B cells by flow sorting (CD20+, IgD-),
 provided by Dr. Louis M. Staudt (NCI), Dr. David Allman
 (NCI) and Dr. Gerald Marti (CBER). cDNA synthesis was
 primed with a Not I - oligo(dT) primer
 [5'-GTGTACCAATCGAATGGAGCGCGCCCTCATTTTCTTTTCTTTT-
 3']. Double-stranded cDNA was ligated to Eco RI adaptors
 (Pharmacia), digested with Not I and cloned into the Not I
 and Eco RI sites of the modified pT73 vector. Library
 went through one round of normalization, and was
 constructed by Bento Soares and M. Fatima Bonaldo."
 /db_xref="taxon:9606"
 /clone_lib="IMAGE:1270592"
 /clone_lib="NCI_CGAP_GCB1"
 /tissue_type="germinal center B cell"
 /lab_host="DH10B"

source

BASE COUNT 68 a 95 c 87 g 65 t

ORIGIN

Query Match 7.1%; Score 285; DB 15; Length 315;
 Best Local Similarity 96.2%; Pred. No. 0.00e+00;
 Matches 302; Conservative 0; Mismatches 11; Indels 1; Gaps 1;

1 TTTTTCCTCAAACTGAAACTCATATATTCAGTATTTTACTCCACAGCACCTCCCCCC 60
 4015 TTTTTCCTCAAACTGAAACTCATATATTCAGTATTTTACTCCACAGCACCTCCCCCC 3956

Db 61 AATTGACCCAGGACCCCAATCCAGGTGCAGGTCCTCGGCTGTGTACAGGGCACAC 120
 Cp 3955 AATTGACCCAGGACCCCAATCCAGGTGCAGGTCCTCGGCTGTGTACAGGGCACAC 3896

Db 121 CTTTGTGCTACTCAATTCACAGCTCCACAGGTCCTCTCAGGGTCTCCACCTGGATG 180
 Cp 3895 CTTTGTGCTACTCAATTCACAGCTCCACAGGTCCTCTCAGGGTCTCCACCTGGATG 3836

Db 181 GTGGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 240
 Cp 3835 GTGGGCGT-GGAGGCGAAGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3777

Db 241 GGACTATTCCTATGTGGGAGTGGAGCCGGGCTCCTGTGTAGGAAAGCTGCCCTGGG 300
 Cp 3776 GGACTATTCCTATGTGGGAGTGGAGCCGGGCTCCTGTGTAGGAAAGCTGCCCTGGG 3717

Db 301 GTGAGCCGAGCGC 314
 Cp 3716 GTGAGCCGAGCGC 3703

RESULT 4 AA311750 409 bp mRNA EST 19-APR-1997
LOCUS EST182469 Jurkat T-cells VI Homo sapiens cDNA 5' end, mRNA
DEFINITION sequence.
ACCESSION AA311750
NID G1964077
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
 Eukaryotes; mitochondrial eukaryotes; Metazoa; Chordata;
 Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea;
 Homo.

REFERENCE
 1 (bases 1 to 409)
 Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A.,
 Bult,C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D.,
 White,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Wai,C.,
 Clayton,R.A., Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D.,
 Fitzgerald,L.M., Fritchugh,W.M., Fritchman,J.L., Geoghagen,N.S.,
 Glodek,A., Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S., Jr.,
 Kelley,J.M., Kelley,J.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M.,
 Moreno-Palauques,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M.,
 Phillips,C.A., Ryder,S.E., Scott,J.L., Saudek,D.M., Shirley,R.,
 Small,K.V., Spriggs,T.A., Utterback,T.R., Weidman,J.F., Li,Y.,
 Bednarek,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J.,
 Dimke,D., Feng,D.-F., Ferrie,A., Fischer,C., Hastings,G.A.,
 He,W.W., Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K.,
 Kozak,D.L., Kunsch,C., Hungjun,J., Li,H., Weissner,P.S., Olsen,H.,
 Raymond,L., Wei,Y.F., Wing,J., Xu,C., Yu,G.L., Ruben,S.M.,
 Dillion,P.J., Fannon,M.R., Rosen,C.A., Haseltine,W.A., Fields,C.,
 Fraser,C.M. and Venter,J.C.
 Initial assessment of human gene diversity and expression patterns
 based upon 83 million nucleotides of cDNA sequence
 Nature 377 (6547 Suppl), 3-174 (1995)
 96026280

TITLE
JOURNAL
MEDLINE
COMMENT

Contact: Kerlavage, AR
 Bioinformatics
 The Institute for Genomic Research
 9712 Medical Center Drive, Rockville, MD 20850 USA
 Tel: 3018699055
 Fax: 3018699423
 Email: arkerlavet@tigr.org
 For clone availability, additional sequence and expression
 information related to this EST, please check the TIGR Human Gene
 Index (<http://www.tigr.org/tdb/hgi/hgi.html>)
 Seq primer: M13 Reverse.

FEATURES
 Location/Qualifiers
 1..409
 /organism="Homo sapiens"
 /note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
 XbaI"
 /db_xref="ARCC (inhost):158964"
 /db_xref="taxon:9606"
 /clone_lib="Jurkat T-cells VI"
 /cell_type="T-lymphocyte"

BASE COUNT 65 a 120 c 133 g 86 t 5 others

ORIGIN

Query Match 6.2%; Score 248; DB 4; Length 409;
 Best Local Similarity 99.2%; Pred. No. 0.00e+00;
 Matches 248; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Db 1 GTTTGGTGATGATTTCTTTGTGTGACACTCACCTCACCCACGCGAAACCTTCCTCA 60
 QY 2649 GTTTGGTGATGATTTCTTTGTGTGACACTCACCTCACCCACGCGAAACCTTCCTCA 2708

Db 61 GGACCTGTGTCGAGGTGTCCTCGTGTGATGCTCGGTGTGAACTTGGGGAAGACAGTGG 120
 QY 2709 GGACCTGTGTCGAGGTGTCCTCGTGTGATGCTCGGTGTGAACTTGGGGAAGACAGTGG 2768

Db 121 TGAACCTTCCCTGTAGAGACGAGGCCCTCGGTGCGACGGCTTTTTCAGATCGCGCCC 180

[illegible]

CDNA clone 97SN1787, mRNA sequence.

AA754459
NID 92801165
KEYWORDS EST.
SOURCE rice.
ORGANISM Oryza sativa

REFERENCE
AUTHORS Nahm,B.H., Kim,J.K., Cheong,J.J., Kim,S.I., Hahn,T.R., Moon,E.P., Kim,W.T., Kim,W.Y., Yang,M.S., Park,R.D., Sohn,U.I., Kang,K.Y., Lee,M.C. and Eun,M.Y.
TITLE Large-scale Sequencing Analysis of ESTs from Rice Immature Seed Unpublished (1998)

COMMENT Contact: Eun M.Y.
Department of CytoGenetics
National Inst. of Agri. Sci. and Tech, RDA
Suwon, KyungGido, Korea
Tel: 82 331 290 0301
Fax: 82 331 290 0307
Email: myeun@sun20.osti.re.kr
Submitted by Baek Hie Nahm, Dept of Biological Science, Myongji University, Yongin, Korea. 449-728 bnhahm@bioserver.myongji.ac.kr
Seq primer: M13 Reverse Primer.
Location/Qualifiers
1. .252
/organism="Oryza sativa"
/cultivar="Milyang23"
/note="Vector: pBluescript SK(+); Site_1: EcoRI; Site_2: XhoI; Directional cDNA library inserted into lambda ZAPII vector at 5' end with EcoRI and 3' end with Xho I site."
/db_xref="taxon:4530"
/clone="97SN1787"
/clone_lib="Rice Immature Seed Lambda ZAPII cDNA Library"
/tissue_type="Immature Seed"
/dev_stage="5 days after pollination"
/lab_host="E. coli SOLR"

BASE COUNT 5 a 21 c 12 g 35 t 179 others
ORIGIN

Query Match 1.5%; Score 62; DB 13; Length 252;
Best Local Similarity 11.7%; Pred. No. 5.26e-57;
Matches 27; Conservative 118; Mismatches 82; Indels 3; Gaps 3;

Db 23 YBCHGNBVTVASHGNYMSVHNCTBRGTHCDCKNNVNWSTWGTGVNBNVSGDHWBYB 82
3021 TGCAGGTGAACACCTCCACAGCGTGTGCACACATCTACAGACATCCCTCGTCGAG 3080
83 BNTKVDVGNHT-RCS-RWRBVTMAHYHDYTNBBYNNNDYHMHBBYBGTGCTCTMW 140
Qy 3081 CGTACAGTTTCCAGCATGTGCTCGAGCTCCATTCATCAGCA-AGTTTGAAGAAG 3139
141 CWBHYNTKCTASGHTSTNYDVKSSTNMGVTSBYSKSMHGYWCSBVBKHYTKVSTTRAT 200
Qy 3140 CCCACATTTTCTGCGCGCATCTCTGACACGCGCTCCCTGCTACTCCATCCCTGAAA 3199
Db 201 RSTCVKRYKVMWTKVVKYHYVBBGCHBTDSKTKTMMNKHYMTST 250
Qy 3200 GCCAAGAACGACGGATGTCGCTGGGGGCCAAGGGCGCCGCCCTCT 3249

RESULT 8
LOCUS AA754458 247 bp mRNA EST 20-JAN-1998
DEFINITION 97SN1784 Rice Immature Seed Lambda ZAPII cDNA Library Oryza sativa
ACCESSION AA754458
NID 92801164
KEYWORDS EST.
SOURCE rice.
ORGANISM Oryza sativa

REFERENCE
AUTHORS Nahm,B.H., Kim,J.K., Cheong,J.J., Kim,S.I., Hahn,T.R., Moon,E.P., Kim,W.T., Kim,W.Y., Yang,M.S., Park,R.D., Sohn,U.I., Kang,K.Y., Lee,M.C. and Eun,M.Y.
TITLE Large-scale Sequencing Analysis of ESTs from Rice Immature Seed Unpublished (1998)

COMMENT Contact: Eun M.Y.
Department of CytoGenetics
National Inst. of Agri. Sci. and Tech, RDA
Suwon, KyungGido, Korea
Tel: 82 331 290 0301
Fax: 82 331 290 0307
Email: myeun@sun20.osti.re.kr
Submitted by Baek Hie Nahm, Dept of Biological Science, Myongji University, Yongin, Korea. 449-728 bnhahm@bioserver.myongji.ac.kr
Seq primer: M13 Reverse Primer.
Location/Qualifiers
1. .252
/organism="Oryza sativa"
/cultivar="Milyang23"
/note="Vector: pBluescript SK(+); Site_1: EcoRI; Site_2: XhoI; Directional cDNA library inserted into lambda ZAPII vector at 5' end with EcoRI and 3' end with Xho I site."
/db_xref="taxon:4530"
/clone="97SN1787"
/clone_lib="Rice Immature Seed Lambda ZAPII cDNA Library"
/tissue_type="Immature Seed"
/dev_stage="5 days after pollination"
/lab_host="E. coli SOLR"

BASE COUNT 5 a 21 c 12 g 35 t 179 others
ORIGIN

Query Match 1.5%; Score 62; DB 13; Length 252;
Best Local Similarity 11.7%; Pred. No. 5.26e-57;
Matches 27; Conservative 118; Mismatches 82; Indels 3; Gaps 3;

Db 23 YBCHGNBVTVASHGNYMSVHNCTBRGTHCDCKNNVNWSTWGTGVNBNVSGDHWBYB 82
3021 TGCAGGTGAACACCTCCACAGCGTGTGCACACATCTACAGACATCCCTCGTCGAG 3080
83 BNTKVDVGNHT-RCS-RWRBVTMAHYHDYTNBBYNNNDYHMHBBYBGTGCTCTMW 140
Qy 3081 CGTACAGTTTCCAGCATGTGCTCGAGCTCCATTCATCAGCA-AGTTTGAAGAAG 3139
141 CWBHYNTKCTASGHTSTNYDVKSSTNMGVTSBYSKSMHGYWCSBVBKHYTKVSTTRAT 200
Qy 3140 CCCACATTTTCTGCGCGCATCTCTGACACGCGCTCCCTGCTACTCCATCCCTGAAA 3199
Db 201 RSTCVKRYKVMWTKVVKYHYVBBGCHBTDSKTKTMMNKHYMTST 250
Qy 3200 GCCAAGAACGACGGATGTCGCTGGGGGCCAAGGGCGCCGCCCTCT 3249

RESULT 8
LOCUS AA754458 247 bp mRNA EST 20-JAN-1998
DEFINITION 97SN1784 Rice Immature Seed Lambda ZAPII cDNA Library Oryza sativa
ACCESSION AA754458
NID 92801164
KEYWORDS EST.
SOURCE rice.
ORGANISM Oryza sativa

REFERENCE
AUTHORS Nahm,B.H., Kim,J.K., Cheong,J.J., Kim,S.I., Hahn,T.R., Moon,E.P., Kim,W.T., Kim,W.Y., Yang,M.S., Park,R.D., Sohn,U.I., Kang,K.Y., Lee,M.C. and Eun,M.Y.
TITLE Large-scale Sequencing Analysis of ESTs from Rice Immature Seed Unpublished (1998)

COMMENT Contact: Eun M.Y.
Department of CytoGenetics
National Inst. of Agri. Sci. and Tech, RDA
Suwon, KyungGido, Korea
Tel: 82 331 290 0301
Fax: 82 331 290 0307
Email: myeun@sun20.osti.re.kr
Submitted by Baek Hie Nahm, Dept of Biological Science, Myongji University, Yongin, Korea. 449-728 bnhahm@bioserver.myongji.ac.kr
Seq primer: M13 Reverse Primer.
Location/Qualifiers
1. .252
/organism="Oryza sativa"
/cultivar="Milyang23"
/note="Vector: pBluescript SK(+); Site_1: EcoRI; Site_2: XhoI; Directional cDNA library inserted into lambda ZAPII vector at 5' end with EcoRI and 3' end with Xho I site."
/db_xref="taxon:4530"
/clone="97SN1787"
/clone_lib="Rice Immature Seed Lambda ZAPII cDNA Library"
/tissue_type="Immature Seed"
/dev_stage="5 days after pollination"
/lab_host="E. coli SOLR"

BASE COUNT 5 a 21 c 12 g 35 t 179 others
ORIGIN

Query Match 1.3%; Score 51; DB 13; Length 247;
Best Local Similarity 10.5%; Pred. No. 2.34e-39;
Matches 23; Conservative 108; Mismatches 86; Indels 2; Gaps 2;

Db 17 BAWNKHHTMMTBWCCVRRVGGTTTNGKNGRTTWTWDCSDNAHCRYPVBYVYARSKVG 76
Qy 1672 GATCCTGGCCAAAGTTCCTGCAGTGTGATGATGATGATGATGATGATGATGATGATG 1731
Db 77 YGTBYYSWNVDTNGTGVGKTTVNVHSGWNNRCSNVYVWVATCDYBHYBDRANHYD 136
Qy 1732 TTTCTTTATGTCAGGAGACCAGCTTCAAGAAGACAGCGCTCTTTTCTACCGGAAG 1791
Db 137 DTRCTNDRGYCN-YTASDNGTSATKRVTYGDKTSDCGGGCKRVKTYGSBBYBRCGVN 195
Qy 1792 TGTCTGGAGCAAGTTGCAAGCATGGAATCAGACA-GCATTGAAGAGGGTGCAGCTGC 1850
Db 196 VRTTSMWTDKSTKVBMSMDMSRRSRVHGRWBNKKRGM 234
Qy 1851 GGGAGCTGTCGGAAGACAGATGATGATGATGATGATGATGATGATGATGATGATG 1889

RESULT 9
LOCUS AA299878 303 bp mRNA EST 18-APR-1997
DEFINITION EST12462 uterus tumor I Homo sapiens cdna 5' end, mRNA sequence.
ACCESSION AA299878
NID g1952209
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE
AUTHORS Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A., Bult,C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D., White,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Wai,C., Clayton,R.A., Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D., Fitzgerald,L.M., Fitzhugh,W.M., Fritchman,J.L., Geochagen,N.S.,

CONTACT: Marra M/Mouse EST Project
WashU-HHMI Mouse EST Project

```

1. .000
/organism="Homo sapiens"
/db xref="taxon:9606"

```


[illegible]

Search completed: Mon Aug 3 13:21:09 1998
Job time : 8704 secs.

/db_xref="taxon:9606"
/clone="2288C3"
/clone_lib="CIT-HSP"
/sex="Male"
/cell_type="Sperm"
59 a 92 c 91 g 75 t

BASE COUNT
ORIGIN

Query Match 0.7%; Score 27; DB 26; Length 317;
Best Local Similarity 86.5%; Pred. No. 6.67e-06;
Matches 32; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Db 231 CCTTCCCGCAGGTGCTCTGCTGAGGAGATGGGGC 267
||||| ||||||| ||||||| ||||||| ||| |||
QY 264 CCTTCCCGCAGGTGCTCTGCTGAGGAGCTGGTGC 300

RESULT 15
LOCUS AA775218 560 bp mRNA EST 05-FEB-1998
DEFINITION ac79e03.s1 Stratagene lung (#937210) Homo sapiens cDNA clone 868828
3' similar to gb:J00194 HLA CLASS II HISTOCOMPATIBILITY ANTIGEN, DR
ALPHA CHAIN (HUMAN);, mRNA sequence.

ACCESSION AA775218
NID 92834552
KEYWORDS EST.
SOURCE human.

ORGANISM Homo sapiens
Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 560)
Hallier, L., Allen, M., Bowles, L., Dubuque, T., Geisel, G., Jost, S.,
Krizman, D., Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M.,
Martin, J., Moore, B., Schellenberg, K., Steptoe, M., Tan, F.,
Theising, B., White, Y., Wylie, T., Waterston, R. and Wilson, R.
WashU-NCI human EST Project
Unpublished (1997)

TITLE
JOURNAL
COMMENT
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LILN; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 344.

FEATURES
source

1..560
/organism="Homo sapiens"
/note="Organ: lung; Vector: pBluescript SK-; Site: 1:
EcoRI; Site 2: XhoI; Cloned unidirectionally. Primer:
Oligo dT: normal lung. Average insert size: 1.0 kb;
Uni-ZAP XR Vector; -5' adaptor sequence: 5' GAATTCGGCAGAG
3' -3' adaptor sequence: 5' CTCGAGTTTTTTTTTTT 3'";
/db_xref="taxon:9606"
/clone="868828"
/clone_lib="Stratagene lung (#937210)"
/sex="male"
/dev_stage="72 years"
/lab_host="SOLR cells (kanamycin resistant)"
173 a 118 c 134 g 135 t

BASE COUNT
ORIGIN

Query Match 0.7%; Score 27; DB 13; Length 560;
Best Local Similarity 82.9%; Pred. No. 6.67e-06;
Matches 34; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Db 466 TGATGCCACACACAGTCAGGCCCGGCGACACACCA 506
||||| ||||||| ||||||| ||||||| ||| |||||
QY 3452 TGATGCCACACACAGTCAGGCCCGGCGACACACCA 3492